

Power Allocation for Energy Harvesting Transmitter with Causal Information

Zhe Wang, Vaneet Aggarwal, Xiaodong Wang

Abstract

We consider power allocation for an access-controlled transmitter with energy harvesting capability based on causal observations of the channel fading state. We assume that the system operates in a time-slotted fashion and the channel gain in each slot is a random variable which is independent across slots. Further, we assume that the transmitter is solely powered by a renewable energy source and the energy harvesting process can practically be predicted. With the additional access control for the transmitter and the maximum power constraint, we formulate the stochastic optimization problem of maximizing the achievable rate as a Markov decision process (MDP) with continuous state. To efficiently solve the problem, we define an approximate value function based on a piecewise linear fit in terms of the battery state. We show that with the approximate value function, the update in each iteration consists of a group of convex problems with a continuous parameter. Moreover, we derive the optimal solution to these convex problems in closed-form. Further, we propose power allocation algorithms for both the finite- and infinite-horizon cases, whose computational complexity is significantly lower than that of the standard discrete MDP method but with improved performance. Extension to the case of a general payoff function and imperfect energy prediction is also considered. Finally, simulation results demonstrate that the proposed algorithms closely approach the optimal performance.

Index Terms

Causal information, energy harvesting, fading channel, Markov decision process, power allocation.

Z. Wang and X. Wang are with the Electrical Engineering Department, Columbia University, New York, NY 10027 (e-mail: {zhewang, wangx}@ee.columbia.edu).

V. Aggarwal is with AT&T Labs-Research, Bedminster, NJ 07921 USA (e-mail: vaneet@research.att.com).

I. INTRODUCTION

The utilization of renewable energy is an important characteristic of the green wireless communication systems [1]. Renewable energy powered transmitters can be deployed for wireless sensor networks or cellular networks, reducing the reliance on traditional batteries and prolonging the transmitter's lifetime [2][3]. However, the fluctuation of the energy harvesting together with the variation of the channel fading brings many challenges to the design of energy-harvesting communication systems [4][5].

Wireless transmission schemes for energy-harvesting transmitters have been investigated by a number of recent works [6][7][8][9]. In order to achieve the optimal throughput, a "shortest path" based energy scheduling algorithm was proposed in [6] for a static channel with finite battery capacity and non-causal energy harvesting state. The authors of [7] discussed an MDP model for the case when the energy harvesting and channel fading are known causally and there is no maximum power constraint. A staircase water-filling algorithm was proposed in [7] for the case when the battery capacity is infinite, and the energy harvesting and fading channel states are known non-causally. With a finite battery capacity and non-causal energy harvesting and fading channel states, a water-filling procedure was studied in [8], and with an additional maximum power constraint a dynamic water-filling algorithm was proposed in [9]. The authors of [10] developed an online approximately optimal algorithm based on Lyapunov optimization, which is designed to maximize a utility function, based on the number of packet transmissions in energy harvesting networks. In [11], using the discrete MDP model, a reinforcement learning based approach was used to optimize the number of packet transmissions without the prior knowledge of the statistics of the energy harvesting process and the channel fading process. The authors of [12] considered a static channel with causal knowledge of the stationary Poisson energy arrival process and gave an MDP-based solution to maximize the average throughput with unconstrained transmission power. On the other hand, the throughput optimization problem with causal information on the energy harvesting state and the fading channel state, and under the maximum power constraint, remains open. In this paper, we will tackle this problem.

Specifically, we first consider the power allocation for an access-controlled transmitter, which is powered by a renewable energy source and equipped with a finite-capacity battery and has a maximum power constraint. The channel fading is assumed to be a random variable in a slot and is independent across different slots. For energy harvesting, we first assume that it can be predicted accurately for the scheduling period, which can be realized in practice [13][14], and then later introduce the prediction error variables.

Furthermore, we assume that a control center can temporarily suspend the transmitter's access due to channel congestion. Such channel access control for the transmitter is modeled as a first-order Markov process. Under the above setting, this paper finds the approximately optimal power allocation for both the finite- and infinite-horizon cases.

To obtain the power allocation, we formulate the stochastic optimization problem as a discrete-time and continuous-state Markov decision process (MDP), with the objective of maximizing the sum of the payoff in the current slot and the discounted expected payoffs in the future slots, where the payoff function is the achievable channel rate. Since the state variables including the battery state and the channel state in the MDP problem are continuous, to avoid the prohibitively high complexity for updating the value function caused by the continuous states, this paper introduces an approximate value function. We show that the approximate value function is concave and non-decreasing in the variable corresponding to the energy stored in the battery, which further enables the approximate value function be updated in closed-form. This is then used to find the approximately optimal solution of the power allocation for both the finite- and infinite-horizon cases.

The proposed algorithms provide approximate solutions, whose performances are lower bounded by the standard discrete MDP method. Also, to obtain the solution, we solve at most $\mathcal{O}(B_{\max}/\delta \cdot C)$ convex optimization problems where B_{\max} is the battery capacity, δ is the approximation precision, and C is the length of horizon for the finite-horizon case or the maximum number of iterations for the infinite-horizon case. In particular, for the infinite-horizon case, given a convergence tolerance α , the α -converged solution can be obtained within $\mathcal{O}(\log_{\gamma} \alpha)$ iterations, where γ is the discount factor.

The remainder of the paper is organized as follows. In Section II, we describe the system model, formulate the energy scheduling problem as a continuous-state MDP problem and define the value function. In Section III, we define an approximate value function and prove that the approximate value function is non-decreasing and concave with respect to the continuous battery state. In Section IV, we derive the optimal closed-form procedure for updating the approximate value function and develop the power allocation algorithms for both finite- and infinite-horizon cases. The proposed algorithms are extended to deal with the model with a general payoff function and imperfect energy prediction in Section V. Section VI provides simulation results and Section VII concludes the paper.

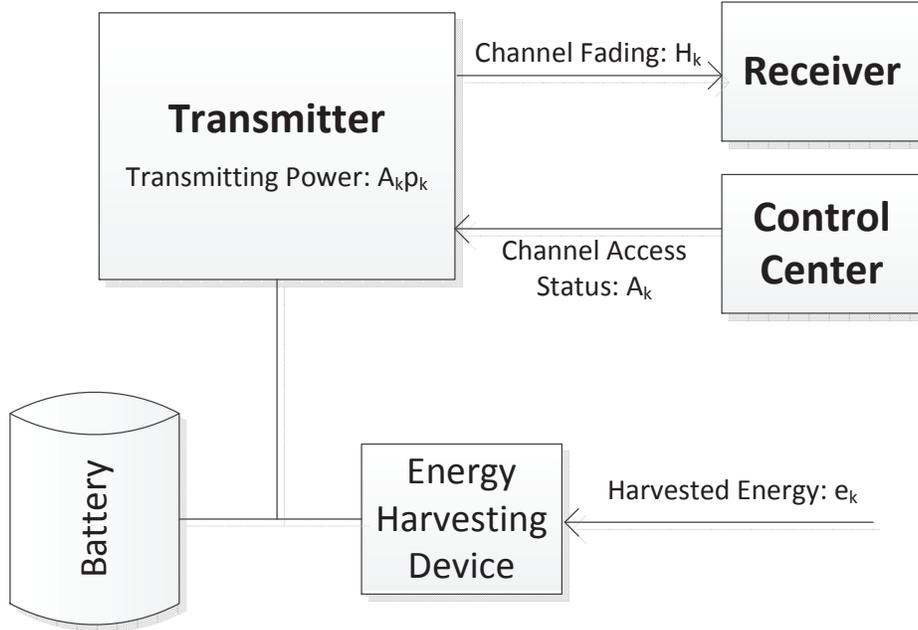


Fig. 1. The system block diagram.

II. PROBLEM FORMULATION

A. System Model

We consider a point-to-point communication system with one transmitter and one receiver, as shown in Fig. 1. We assume a slow fading channel model where the channel gain is constant for a coherence time of T_c (corresponding to a time slot) and changes independently across slots. The signal model for slot k is given by

$$\mathbf{y}_k = H_k \mathbf{x}_k + \mathbf{w}_k, \quad (1)$$

where $\mathbf{y}_k \in \mathbb{C}^{T_c}$ is the received signal, $\mathbf{x}_k \in \mathbb{C}^{T_c}$ is the transmitted signal, $H_k \in \mathbb{C}$ is the channel gain in slot k and $\mathbf{w}_k \in \mathbb{C}^{T_c}$ is the additive white Gaussian noise consisting of $\mathcal{CN}(0, 1)$ elements.

At the beginning of each slot, the transmitter is informed of the channel access status $A_k \in \{0, 1\}$ for the current slot from the control center, where $A_k = 0$ indicates that the channel access is not permitted for slot k while $A_k = 1$ indicates otherwise. We assume that A_k follows a stationary first-order Markov process, whose transition probabilities are given as $\Pr(A_{k+1} = 0 \mid A_k = 1) = q_k$ and $\Pr(A_{k+1} = 0 \mid A_k = 0) = \tilde{q}_k$. If $A_k = 0$, then the transmit power in slot k is $p_k = 0$. On the other hand, if $A_k = 1$, then the transmitter needs to decide its transmit power p_k .

The transmitter is powered by an energy harvesting device, e.g., a solar panel, and a battery. The battery,

which buffers the harvested energy, has a finite capacity, denoted by b_{\max} . Since the energy harvesting process is steady or can be well predicted, we assume that the energy harvested over the next K slots can be non-causally known, denoted as e_k (the causal energy harvesting model will be considered in Section V). We assume $h_k \triangleq |H_k|^2$ is independent across slots (i.i.d. when $K = \infty$).

In slot k , the transmitter transmits at a power level of p_k ($p_k = 0$ if $A_k = 0$), which is constrained by the maximum transmission power p_{\max} and the available energy b_k , i.e.,

$$0 \leq p_k \leq \min \{p_{\max}, b_k/T_c\} . \quad (2)$$

The battery level at the beginning of slot $k + 1$ is given as

$$b_{k+1} = \min \{b_{\max}, b_k + e_k - p_k T_c\} , \quad (3)$$

with the constraint that the battery level is non-negative for all slots, i.e.,

$$b_k \geq 0 . \quad (4)$$

Further, the transmitter receives a payoff $r(p, h)$ based on the transmission power and channel gain. In this paper, we use the achievable channel rate as the payoff, i.e., $r(p, h) = \log(1 + ph)$. Also, in Section V, we consider a general payoff function $r(p, h)$ which is continuous, non-decreasing, and concave with respect to p given h .

B. Problem Formulation

We assume that e_k can be predicted non-causally while all other variables are only known causally to the transmitter (we will relax this assumption in Section V where we assume that e_k is predicted with a random error ε_k). Denote $\mathbf{H} \triangleq [h_1, h_2, \dots, h_K]$, $\mathbf{A} \triangleq [A_1, A_2, \dots, A_K]$, and a discount factor $\gamma \in [0, 1]$. We assume that all the side information, e.g., the distributions of all random variables and the predictions of the harvested energy, is known before the first slot. Then the power allocation policy $\mathcal{P} \triangleq \{p_k(\Gamma_k) \mid k = 1, 2, \dots, K\}$ needs to be calculated to maximize the expected total payoff in the next K slots, where $\Gamma_k \triangleq (b_k, h_k, A_k)$ consists of the observations available at the beginning of slot k . Since b_k and h_k are continuous variables, it is not possible to store \mathcal{P} in a look-up table. Instead, we only store some of the intermediate results, i.e., the approximate value function introduced in Section III, in an efficient way, and then calculate the power allocation when Γ_k is observed. Specifically, at the beginning

of slot k , given Γ_k , if channel access is permitted, i.e., $A_k = 1$, the transmitter calculate the power level p_k . And if the channel access is not permitted, i.e., $A_k = 0$, then $p_k = 0$. To that end, we formulate the following optimization problem for defining the optimal policy

$$\mathcal{P}^* \triangleq \arg \max_{p_k(\cdot), k=1,2,\dots,K} \left\{ \mathbb{E}_{\mathbf{H}, \mathbf{A}} \left[\sum_{k=1}^K \gamma^{k-1} \log(1 + p_k(\Gamma_k) h_k) \right] \right\}, \quad (5)$$

subject to the constraints in (2), (3), and (4) for $k = 1, 2, \dots, K$.

Note that by (3), the battery level b_k forms a continuous-state first-order Markov chain, whereas the channel access state A_k is a discrete-state Markov chain by assumption. Then, we can convert the problem in (5) to its equivalent MDP recursive form [15] in terms of the *value function*, which represents the total payoff received in the current slot and expected to be received in the future slots.

Specifically, in the MDP model we treat the battery level b and the channel access state A , i.e., (b, A) , as the state, the channel h as the observation, and the transmit power p as the decision. Then, the state space becomes $\{0 \leq b \leq b_{\max}\} \times \{0, 1\}$; and the corresponding decision space is $\mathcal{D}_1(b) = \{0 \leq p \leq \min\{b/T_c, p_{\max}\}\}$ and $\mathcal{D}_0 = \{0\}$, corresponding to $A = 1$ and $A = 0$, respectively. The value function is then recursively defined as

$$v^k(b_k, A_k) \triangleq \mathbb{E}_{h_k} \left[\max_{p_k(\Gamma_k) \in \mathcal{D}_{A_k}(b_k)} \left\{ \log(1 + p_k(\Gamma_k) h_k) + \gamma u^k(b_k, p_k(\Gamma_k), A_k) \right\} \right], \quad k = 1, 2, \dots, K, \quad (6)$$

where

$$u^k(b_k, p_k, A_k) \triangleq \mathbb{E}_{A_{k+1}|A_k} \left[v^{k+1}(\min\{b_{\max}, b_k + e_k - p_k T_c\}, A_{k+1}) \right], \quad (7)$$

and

$$v^{K+1}(b, A) = 0, \quad \text{for all } b \in [0, b_{\max}], A \in \{0, 1\}. \quad (8)$$

Note that, $v^k(b_k, A_k)$ represents the expected maximum discounted payoff between slots k and K given the side information b_k and A_k . Due to the causality and the backward recursion, the observation Γ_k in slot k does not affect the value function for slot $k + 1$. Also, when $A_k = 1$, given the value function for slot $k + 1$, the optimal power allocation for slot k can be obtained by

$$p_k^*(\Gamma_k) = \arg \max_{p \in \mathcal{D}_{A_k}(b_k)} \left\{ \log(1 + p h_k) + \gamma u^k(b_k, p, 1) \right\}, \quad (9)$$

where $u^k(b, p, A)$ is calculated using (7). Moreover, when $A_k = 0$, we always have

$$p_k^*(\Gamma_k) = 0 . \quad (10)$$

III. APPROXIMATE VALUE FUNCTION

By recursively computing the value function $v^k(b, A)$ defined in (6), in theory we can obtain the optimal solution to (9) for each $k \in \{1, 2, \dots, K\}$. However, a closed-form expression for $v^k(b, A)$ is hard to obtain when K is large, e.g., $K \geq 3$. A typical approach is to quantize the continuous variables (b, p, h) to finite number of discrete levels, i.e., to convert the original problem to a discrete MDP problem [15]. However, with such discretization, solving the corresponding discrete MDP problem involves an exhaustive search on $\mathcal{D}_1(b)$ for all discretized h , and we can only obtain discrete power levels.

In order to efficiently solve the MDP problem and obtain the continuous power allocation, in this section, we will define an approximate value function by using a piecewise linear approximation based on some discrete samples of $\{v^k(B, A) \mid B \in \{0, \delta, 2\delta, \dots, b_{\max}\}, A \in \{0, 1\}\}$ where δ is an approximation precision. This approximate value function is shown to be concave and non-decreasing in the variable corresponding to the energy stored in the battery, making the optimal power allocation problem in (9) (or (18)) a convex optimization problem.

A. Value Function Approximation

With an approximation precision parameter δ , we define a piecewise linear approximation operator:

$$\mathcal{L} [v^k(b, A), \delta] \triangleq v^k(\lfloor b/\delta \rfloor \delta, A) + \frac{b - \lfloor b/\delta \rfloor \delta}{\delta} (v^k(\lceil b/\delta \rceil \delta, A) - v^k(\lfloor b/\delta \rfloor \delta, A)), \quad b \in [0, b_{\max}] , \quad (11)$$

and $\mathcal{L} [v^K(b, A), \delta] \triangleq v(b_{\max}, A)$ for any $b > b_{\max}$, as shown in Fig. 3.

Initially, we define

$$W_\delta^K(b, A) \triangleq \mathcal{L} [v^K(b, A), \delta] , \quad (12)$$

which is a linear approximation to $v^K(b, A)$. Then, recursively from $k = K - 1$ to $k = 1$, we use the approximate value function to replace the original value function in (7), i.e., $v^k(b, A) \leftarrow W_\delta^k(b, A)$, and define

$$U^k(b_k, p_k, A_k) \triangleq \mathbb{E}_{A_{k+1}|A_k} [W_\delta^{k+1}(\min\{b_{\max}, b_k + e_k - p_k T_c\}, A_{k+1})] . \quad (13)$$

By setting $u^k(b_k, p_k, A_k) \leftarrow U^k(b_k, p_k, A_k)$ in (6), we further define

$$V^k(b_k, A_k) \triangleq \mathbb{E}_{h_k} \left[\max_{p_k(\Gamma_k) \in \mathcal{D}_{A_k}(b_k)} \left\{ \log(1 + p_k(\Gamma_k)h_k) + \gamma U^k(b_k, p_k(\Gamma_k), A_k) \right\} \right]. \quad (14)$$

Finally, we write the approximation value function as

$$W_\delta^k(b, A) \triangleq \mathcal{L} [V^k(b, A), \delta]. \quad (15)$$

Note that, in (13)-(15), we made the substitutions $v^k(b, A) \leftarrow W_\delta^k(b, A)$ and $u^k(b_k, p_k, A_k)$ in (7) and (6), respectively. Thus we can treat the approximate value function $W_\delta^k(b, A) \triangleq \mathcal{L} [V^k(b, A), \delta]$, which is updated by (13)-(15), as an approximation to the value function $v^k(b, A)$, which is updated by (6)-(7).

We consider the approximation error $\|W_\delta^k(b, A) - v^k(b, A)\|_\infty$ at slot k (or iteration $i = K - k + 1$). In each iteration, the error is produced by the piecewise linear approximation in (15) and propagated through solving the problem in (14). Then, at the end of each iteration the total error accumulated by the obtained approximate value function is the sum of the newly produced error and the discounted propagated error, growing with the iteration number. Since the update rules for both $v^k(b, A)$ and $W_\delta^k(b, A)$ start from the same initial value function $v^K(b, A)$, then the total error in the i -th iteration (we use the subscript (i) to denote the i -th iteration, which represents slot $K - i + 1$) can be bounded by

$$\|W_\delta^{(i)}(b, A) - v^{(i)}(b, A)\|_{\max} \leq \sum_{j=1}^i \gamma^{i-j} \epsilon_j(\delta) \quad (16)$$

where

$$\epsilon_j(\delta) \triangleq \max_{b \in [0, b_{\max}], A \in \{0, 1\}} \{V^{(j)}(b, A) - W_\delta^{(j)}(b, A)\} = \|V^{(j)}(b, A) - W_\delta^{(j)}(b, A)\|_\infty \quad (17)$$

is the new error produced by (15) in the j -th iteration.

With the approximate value function for each slot k , when $A = 1$, the power allocation given Γ can be obtained by

$$p_k^*(\Gamma) = \arg \max_{p \in \mathcal{D}_1(b)} \left\{ \log(1 + ph) + \gamma U^k(b, p, 1) \right\}. \quad (18)$$

Define $\mathcal{B}_\delta \triangleq \{0, \delta, 2\delta, \dots, b_{\max}\}$. Note that the approximate value function is linearly recovered from the sample set $\{V^k(b, A) \mid b \in \mathcal{B}_\delta\}$ and $W_\delta^k(b, A) = V^k(b, A)$ for all $b \in \mathcal{B}_\delta$. We can consider the standard dynamic programming with the discretized state space as a special case of the update rules in (13)-(15). Then, the performance achieved with the approximate value function can be characterized as follows.

Proposition 1: The approximate value function obtained by recursively solving (13)-(15) is no less than the discrete value function obtained by the standard dynamic programming method with the state space $\mathcal{B}_\delta \times \{0, 1\}$ where δ is the approximate precision.

Proof: Given the discrete state space $\mathcal{B}_\delta \times \{0, 1\}$, since $W_\delta^{(i)}(B, A) = V^{(i)}(B, A)$ for any $B \times A \in \mathcal{B}_\delta \times \{0, 1\}$, the standard dynamic programming follows the same update rule in (13)-(15) but with a discrete feasible power allocation set for the optimization problem in (14), which is a subset of $D_1(b)$. ■

Moreover, in the standard discrete dynamic programming, we discretize all continuous variables, i.e., b_k, h_k, e_k, p_k , and then perform the dynamic programming with an exhaustive search on p_k for all possible combinations of (b_k, h_k) ; while with the proposed approximate value function, we only discretize the battery state b_k and then obtain the approximate value function for each discretized b_k in closed-form.

B. Concavity of Approximate Value Function

In (13)-(15), we note that the approximate value function is based on the solution to an optimization problem (14). To facilitate solving (14), in this subsection, we will show that the approximate value function $W_\delta^k(b, A)$ given in (15) is concave for $0 \leq b \leq b_{\max}$ given $A \in \{0, 1\}$. Then (14) is a convex optimization problem given h and b .

First, we introduce the following lemma, which can be easily shown and illustrated in Fig. 2.

Lemma 1: If a function $f(x) \in \mathbb{R}$ ($x \in \mathcal{X} \subseteq \mathbb{R}$) is non-decreasing, for any $x' \in \mathcal{X}$, $f(\min\{x, x'\})$ is also non-decreasing. Further, if the non-decreasing function $f(x)$ is concave, then $f(\min\{x, x'\})$ is concave for $x \in \mathcal{X} \cup [x', \infty)$.

We have the following non-decreasing property of $W_\delta^k(b, A)$.

Proposition 2: For any $k \in \{1, 2, \dots, K-1\}$, if the approximate value function $W_\delta^{k+1}(b, A)$ is non-decreasing with respect to $b \in [0, b_{\max}]$ given $A \in \{0, 1\}$, so is $W_\delta^k(b, A)$.

Proof: If $W_\delta^{k+1}(b, A)$ is non-decreasing with respect to $b \in [0, b_{\max}]$ for $A \in \{0, 1\}$, by Lemma 1, we have that $W_\delta^{k+1}(\min\{b_{\max}, b\}, A)$ is also non-decreasing with respect to $b \in [0, +\infty)$. Then, we have that $U^k(b, p, A)$, which is a linear combination of the terms of the form $W_\delta^{k+1}(\min\{b_{\max}, b + e_k - p_k T_c\}, A)$, is also non-decreasing with respect to $b \in [0, b_{\max}]$, given p and A .

Given any battery level $b \in [0, b_{\max})$, channel fading h , the power p_0 such that $p_0 \in \mathcal{D}_A(b)$, and $\epsilon > 0$ such that $b + \epsilon \leq b_{\max}$, we have

$$p_0 \in \mathcal{D}_A(b + \epsilon), \quad (19)$$

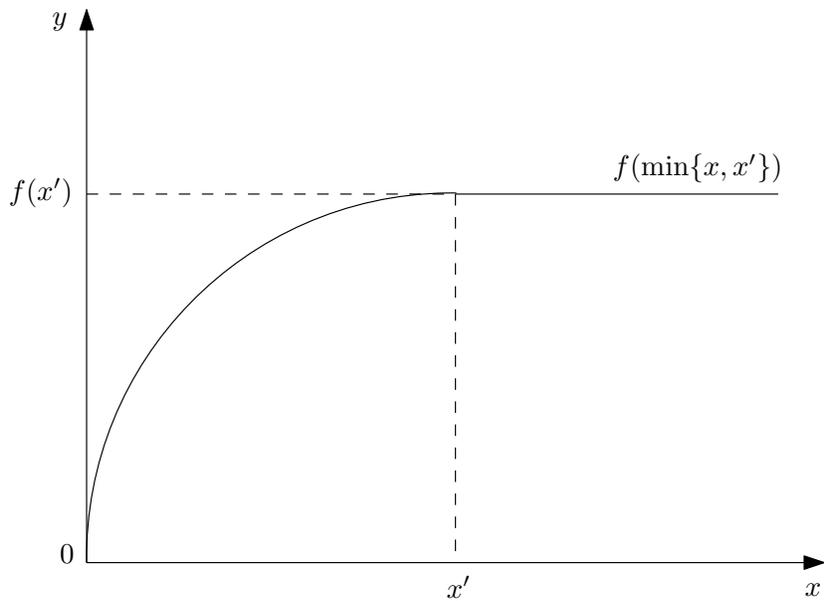


Fig. 2. Illustration of Lemma 1.

and

$$\log(1 + p_0 h) + \gamma U^k(b, p_0, A) \leq \log(1 + p_0 h) + \gamma U^k(b + \epsilon, p_0, A) \quad (20)$$

$$\leq \max_{p \in \mathcal{D}_A(b+\epsilon)} \{ \log(1 + ph) + \gamma U^k(b + \epsilon, p, A) \}. \quad (21)$$

Since $V^k(b, A)$ is a non-negative linear combination of the terms of the form $\max_{p \in \mathcal{D}_A(b)} \{ \log(1 + ph) + U^k(b, p, A) \}$, $V^k(b, A)$ is non-decreasing with respect to $b \in [0, b_{\max}]$. Then, by (15), we have that $W_\delta^k(b, A)$ is also non-decreasing with respect to $b \in [0, b_{\max}]$. ■

The next result is on the concavity of $W_\delta^k(b, A)$.

Proposition 3: For any $k \in \{1, 2, \dots, K\}$, if the approximate value function $W_\delta^{k+1}(b, A)$ is non-decreasing and concave with respect to $b \in [0, b_{\max}]$ given $A \in \{0, 1\}$, so is $W_\delta^k(b, A)$.

Proof: Since $W_\delta^{k+1}(b, A)$ is non-decreasing and concave with respect to $b \in [0, b_{\max}]$ given $A \in \{0, 1\}$, by Lemma 1, we have $W_\delta^{k+1}(\min\{b_{\max}, b\}, A)$ is non-decreasing and concave with respect to $b \geq 0$ given $A \in \{0, 1\}$. Since $b + e - pT_c$ is a linear combination of b and p , then $W_\delta^{k+1}(\min\{b_{\max}, b + e - pT_c\}, A)$ is jointly concave with respect to b and p . Moreover, it follows that $U^k(b, p, A)$ is also jointly concave with respect to b and p given $A \in \{0, 1\}$ [16].

Since the feasible domain $\mathcal{D}_A(b)$ is different under $A = 0$ and $A = 1$. We consider the two cases separately.

When $A = 0$, since $\mathcal{D}_0 = 0$, $v^k(b, 0)$ can be written as

$$V^k(b, 0) = \mathbb{E}_{h_k} \left[\gamma U^k(b, 0, 0) \right]. \quad (22)$$

Since $U^k(b, p, A)$ is concave with respect to $b \in [0, b_{\max}]$ given p and $A \in \{0, 1\}$, so is $V^k(b, 0)$ [16].

Then, by (15), $W_\delta^k(b, 0)$ is non-decreasing with respect to $b \in [0, b_{\max}]$.

When $A = 1$, the feasible domain of the objective function in (6) is given by $\mathcal{C} \triangleq \{(b, p) : 0 \leq b \leq b_{\max}, 0 \leq p \leq \min\{b/T_c, p_{\max}\}\}$. It can be verified that \mathcal{C} is a convex set. Then, for any $(b_1, p_1), (b_2, p_2) \in \mathcal{C}$, their convex combination $(\theta b_1 + \bar{\theta} b_2, \theta p_1 + \bar{\theta} p_2) \in \mathcal{C}$, where $\theta \in [0, 1]$ and $\bar{\theta} \triangleq 1 - \theta$.

Moreover, since $\mathcal{D}_1(b_1), \mathcal{D}_1(b_2)$ are non-empty, we can denote

$$p_1 = \arg \max_{p \in \mathcal{D}_1(b_1)} \left\{ \log(1 + ph) + \gamma U^k(b_1, p, 1) \right\}, \quad (23)$$

and

$$p_2 = \arg \max_{p \in \mathcal{D}_1(b_2)} \left\{ \log(1 + ph) + \gamma U^k(b_2, p, 1) \right\}. \quad (24)$$

Then

$$\begin{aligned} & \max_{p \in \mathcal{D}_1(\theta b_1 + \bar{\theta} b_2)} \left\{ \log(1 + ph) + \gamma U^{k+1}(\theta b_1 + \bar{\theta} b_2, p, 1) \right\} \\ & \leq \log(1 + (\theta p_1 + \bar{\theta} p_2)h) + \gamma U^{k+1}(\theta b_1 + \bar{\theta} b_2, \theta p_1 + \bar{\theta} p_2, 1) \\ & \leq \theta \log(1 + p_1 h) + \bar{\theta} \log(1 + p_2 h) + \theta \gamma U^{k+1}(b_1, p_1, 1) + \bar{\theta} \gamma U^{k+1}(b_2, p_2, 1) \end{aligned} \quad (25)$$

$$\begin{aligned} & = \theta (\log(1 + p_1 h) + \gamma U^{k+1}(b_1, p_1, 1)) + \bar{\theta} (\log(1 + p_2 h) + \gamma U^{k+1}(b_2, p_2, 1)) \\ & = \theta \max_{p \in \mathcal{D}_1(b_1)} \left\{ \log(1 + ph) + \gamma U^{k+1}(b_1, p, 1) \right\} + \bar{\theta} \max_{p \in \mathcal{D}_1(b_2)} \left\{ \log(1 + ph) + \gamma U^{k+1}(b_2, p, 1) \right\}, \end{aligned} \quad (26)$$

where (25) follows from the joint concavity, and (26) follows from the definitions in (23) and (24).

Therefore, we have that $\max_{p \in \mathcal{D}_1(b)} \left\{ \log(1 + ph) + \gamma U^{k+1}(b, p, 1) \right\}$ is concave with respect to $b \in [0, b_{\max}]$. By (14) and (15), we further have $W_\delta^k(b, 1)$ is concave with respect to $b \in [0, b_{\max}]$ [16]. ■

From Propositions 2 and 3, we have that if $W_\delta^{k+1}(b, A)$ is non-decreasing and concave so is $W_\delta^k(b, A)$ for any $k \in \{1, 2, \dots, K-1\}$. Since $\log(1 + ph)$ is non-decreasing and concave with respect to $b \in [0, b_{\max}]$, it is easily verified by (6) that $W_\delta^K(b, A) = V^K(b, A) = v^K(b, A)$ is also non-decreasing and concave with respect to $b \in [0, b_{\max}]$ given A . By induction, we obtain the following theorem.

Theorem 1: For $k = 1, 2, \dots, K$, the approximate value function $W_\delta^k(b, A)$ is non-decreasing and

concave with respect to $b \in [0, b_{\max}]$ given $A \in \{0, 1\}$. Further, the problem in (14) is a convex optimization problem given $b \in [0, b_{\max}]$ and $A \in \{0, 1\}$.

Since both $V^{(i)}(b, A)$ and $W_\delta^{(i)}(b, A)$ are concave and non-decreasing, where $i = K - k + 1$ is the iteration number, we can further bound the approximation error $\epsilon_i(\delta)$ in (17) as follows.

Proposition 4: For any iteration i , given A , we have

$$0 \leq \epsilon_i(\delta) \leq 2V^{(i)}(\delta, A) - V^{(i)}(2\delta, A) - V^{(i)}(0, A). \quad (27)$$

Proof: By Theorem 1, $V^{(i)}(b, A)$ is non-decreasing and concave with respect to b given A . As illustrated in Fig. 3, for $b \in [0, \delta]$, the value of $V^{(i)}(b, A)$ is smaller than the value on line (*) but larger than $W_\delta^{(i)}(b, A)$, and therefore the distance between the value on line (*) and $W_\delta^{(i)}(b, A)$ can also be considered as an upper bound on the approximation error, i.e., $V^{(i)}(b, A) - W_\delta^{(i)}(b, A)$ for $b \in [0, \delta]$. According to the second-order derivative property of the concave function, we have that

$$\begin{aligned} & V^{(i)}((n+1)\delta, A) - V^{(i)}(n\delta, A) - (V^{(i)}((n+2)\delta, A) - V^{(i)}((n+1)\delta, A)) \\ & \geq V^{(i)}((n+2)\delta, A) - V^{(i)}((n+1)\delta, A) - (V^{(i)}((n+3)\delta, A) - V^{(i)}((n+2)\delta, A)) \end{aligned} \quad (28)$$

for all $n \geq 0$. Then, we further have that $0 \leq \epsilon_i(\delta) \leq \max\{2V^{(i)}(\delta, A) - V^{(i)}(2\delta, A) - V^{(i)}(0, A), 2V^{(i)}(2\delta, A) - V^{(i)}(3\delta, A) - V^{(i)}(\delta, A), \dots\} = 2V^{(i)}(\delta, A) - V^{(i)}(2\delta, A) - V^{(i)}(0, A)$, where $\epsilon_i(\delta) = \|V^{(i)}(b, A) - W_\delta^{(i)}(b, A)\|_\infty$. ■

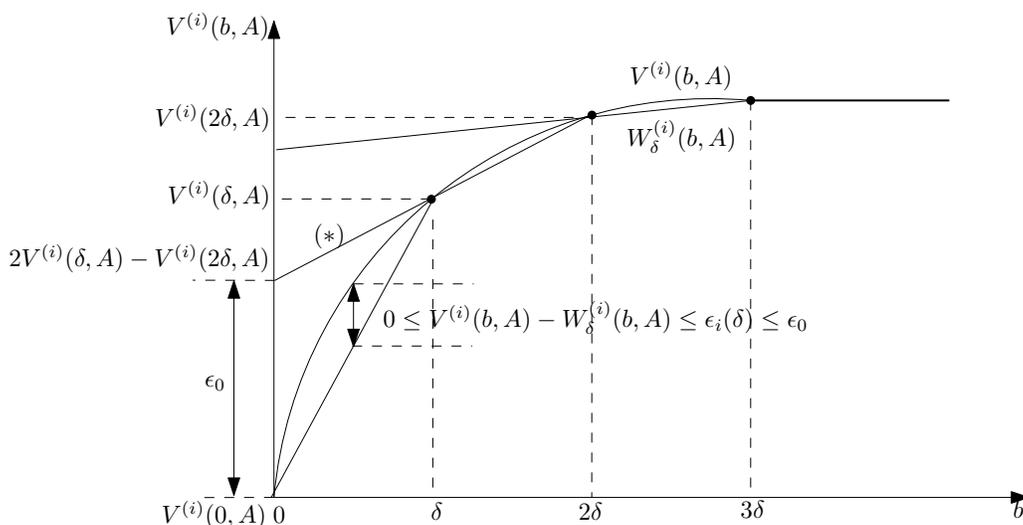


Fig. 3. The piecewise linear approximation of the value function and the approximation error bound.

IV. POWER ALLOCATION WITH PERFECT ENERGY PREDICTION

Note that in (14), we need to solve the following optimization problem for a given $B \in \mathcal{B}_\delta$ and $A \in \{0, 1\}$:

$$p^*(h) = \arg \max_{p(h) \in \mathcal{D}_A(B)} \{ \log(1 + p(h)h) + \gamma U^k(B, p(h), A) \}, \quad h \geq 0. \quad (29)$$

When $A = 0$, $p^*(h) = 0$. On the other hand, when $A = 1$, we will obtain the optimal solution $p^*(h)$ in closed-form.

Since the approximate value function $W_\delta^{k+1}(b, A)$ in (15) is a piecewise linear function of b given A , it follows that $U^k(B, p, 1)$ in (13) is also a piecewise linear function with respect to p given B , which is differentiable everywhere except at $\mathcal{J} \triangleq \{p \mid p = (B + e_k - B_0)/T_c, B_0 \in \mathcal{B}_\delta\}$. By Theorem 1 and Lemma 1, $U^k(B, p, 1)$ is also concave and non-decreasing with respect to p .

Since $U^k(B, p, 1)$ is a piecewise linear function, we denote $\mathcal{I} \triangleq \{p_0, p_1, \dots, p_N\}$ as the set of the non-differentiable points, where $p_0 = 0$, $p_N = \min\{p_{\max}, B/T_c\}$, and p_i , ($0 < i < N$) is the i -th smallest element in $\mathcal{J} \cap \mathcal{D}_1(B) \setminus \{p_0, p_N\}$. Also, we denote $\mathcal{W} = \{w_1, w_2, \dots, w_N\}$ as the set of the corresponding slopes, where w_i is the slope of the segment $[p_{i-1}, p_i]$, given by

$$w_i \triangleq -\frac{\gamma T_c}{\delta} \mathbb{E}_{A|1} \left\{ V^{k+1}(\lceil \min\{b_{\max}, B + e_k - p_i T_c\} / \delta \rceil \delta, A) - V^{k+1}(\lfloor \min\{b_{\max}, B + e_k - p_i T_c\} / \delta \rfloor \delta, A) \right\}, \quad (30)$$

which is derived from (13) and (15). Hence, the derivative of $U^k(B, p, 1)$ for $p \in \mathcal{D}_1(B) \setminus \mathcal{I}$ is

$$w(p) = w_i, \quad \text{if } p \in (p_{i-1}, p_i). \quad (31)$$

Since $U^k(b, p, A)$ is concave and non-decreasing with respect to p , we have $0 \geq w_0 > w_1 > \dots > w_N$. Fig. 4 is a sketch of the stair-case function $w(p)$.

In this section we first obtain the closed-form solution to (29), and then use it to obtain the optimal power allocation for both finite- and infinite-horizon cases.

A. The Optimal Solution to (29)

In this subsection, for simplicity, we drop the superscript k and denote the objective function in (29) as

$$g_h(p) \triangleq \log(1 + ph) + \gamma U(B, p, 1), \quad p \in \mathcal{D}_1(B). \quad (32)$$

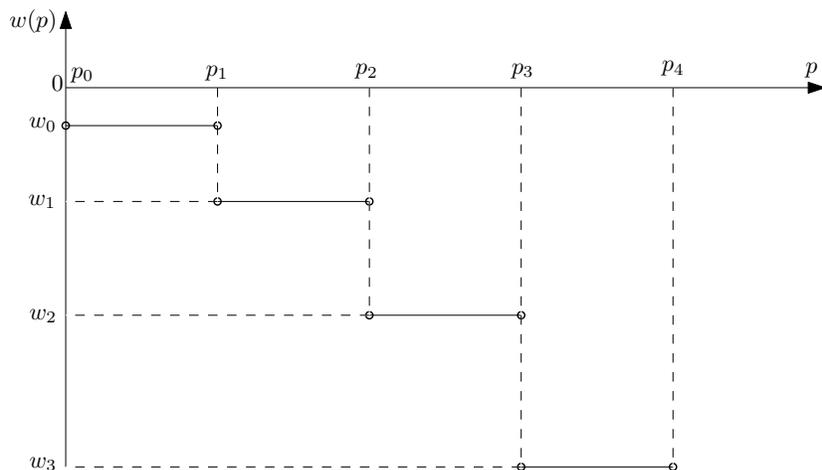


Fig. 4. The derivative of $U^k(B, p, 1)$ with respect to p .

We note that $g_h(p)$ is differentiable for $p \in \mathcal{D}_1(B) \setminus \mathcal{I}$ with

$$g'_h(p) = \frac{1}{1/h + p} + w(p) . \quad (33)$$

On the other hand, at the non-differentiable points in \mathcal{I} , the right-derivative and the left-derivative of $g_h(p)$ can be written as

$$g'_h(p^+) \triangleq \frac{1}{1/h + p} + w(p^+) , \quad (34)$$

and

$$g'_h(p^-) \triangleq \frac{1}{1/h + p} + w(p^-) , \quad (35)$$

respectively.

Theorem 2: The optimal solution to (29) is given by

$$p^*(h) = \begin{cases} -\frac{1}{w_i} - \frac{1}{h} & \frac{1}{h} \in [-\frac{1}{w_i} - p_i, -\frac{1}{w_i} - p_{i-1}] \cap [0, +\infty), i = 1, 2, \dots, N-1 \\ p_i & \frac{1}{h} \in (-\frac{1}{w_{i+1}} - p_i, -\frac{1}{w_i} - p_i) \cap [0, +\infty), i = 1, 2, \dots, N-1 \\ 0 & \frac{1}{h} \in (-\frac{1}{w_1} - p_0, \infty) \\ p_N & \frac{1}{h} \in [0, -\frac{1}{w_N} - p_N) \end{cases} , \quad (36)$$

where $p_0 = 0$ and $p_N = \min\{p_{\max}, B/T_c\}$.

In Fig. 5 we give a sketch of $p^*(h)$. To prove Theorem 2, we first give the necessary and sufficient conditions for the optimal solution p^* as follows [16].

Lemma 2: p^* is the optimal solution to (29) given h , if and only if,

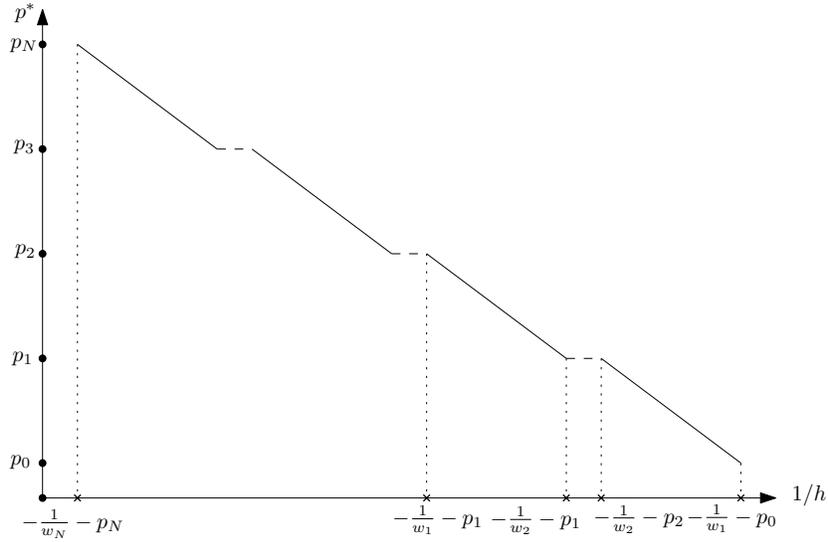


Fig. 5. The optimal solution $p^*(h)$.

- 1) $g'_h(p^{*+}) \leq 0 \leq g'_h(p^{*-})$, when $g'_h(0^+) > 0$ and $g'_h(\min\{B/T_c, p_{\max}\}^-) < 0$;
- 2) $p^* = \min\{B/T_c, p_{\max}\}$, when $g'_h(\min\{B/T_c, p_{\max}\}^-) \geq 0$;
- 3) $p^* = 0$, when $g'_h(0^+) \leq 0$.

Note that, Condition 1 corresponds to the case that p^* is in the interior of $\mathcal{D}_1(B)$. In this case, the left-derivative and the right-derivative should have opposite signs or be both zero at p^* so that the increasing and decreasing of p both lead to the decreasing of the objective function. Condition 2 and Condition 3 correspond to the cases that p^* is on each side of the boundary of $\mathcal{D}_1(B)$, where the objective function is non-decreasing and non-increasing for all $p \in \mathcal{D}_1(B)$, respectively.

The following proposition gives a sufficient condition for the optimality of $p^*(h)$ given B .

Proposition 5: Given any $B \in \mathcal{B}_\delta$, for $h \geq 0$, if the energy schedule $p^*(h) \in \text{int}\mathcal{D}_1(B)$ satisfies

$$p^*(h) = \begin{cases} -\frac{1}{w(p^*(h))} - \frac{1}{h}, & \text{when } p^*(h) \in \text{int}\mathcal{D}_1(B) \setminus \mathcal{I}, \\ -\frac{1}{w(p^*(h)^-)} - \frac{1}{h} \text{ or } -\frac{1}{w(p^*(h)^+)} - \frac{1}{h}, & \text{when } p^*(h) \in \mathcal{I}, \end{cases} \quad (37)$$

then $p^*(h)$ is the optimal solution to (29).

Proof: Substituting (37) into (34)-(35), we have $g'_h(p^*(h)^+) = 0$ or $g'_h(p^*(h)^-) = 0$ when $p^*(h) \in \mathcal{I}$, and $g'_h(p^*(h)^+) = g'_h(p^*(h)^-) = 0$ when $p^*(h) \in \text{int}\mathcal{D}_1(B) \setminus \mathcal{I}$. Since $g'_h(p^*(h)^+) \leq g'_h(p^*(h)^-)$, we have $g'_h(p^*(h)^+) \leq 0 \leq g'_h(p^*(h)^-)$. Moreover, since $g_h(p)$ is concave, we have $0 \leq g'_h(p^*(h)^-) < g'_h(0^-)$ and $g'_h(\min\{p_{\max}, B/T_c\}^-) < g'_h(p^*(h)^+) \leq 0$. By Lemma 2 (Condition 1), we conclude the optimality. ■

Then it is easy to verify that for $\frac{1}{h} \in [-\frac{1}{w_i} - p_i, -\frac{1}{w_i} - p_{i-1}] \cap [0, +\infty)$, $i = 1, 2, \dots, N-1$, the solution given by (36) satisfies the optimality condition in Proposition 5.

For $\frac{1}{h} \in (-\frac{1}{w_{i+1}} - p_i, -\frac{1}{w_i} - p_i) \cap [0, +\infty)$, $i = 1, 2, \dots, N-1$, we use the next proposition to prove the optimality of (36).

Proposition 6: For any non-differentiable point $p_i \in \mathcal{I} \setminus \{p_0, p_N\}$, p_i is the optimal solution to (29) for any $\frac{1}{h} \in (-\frac{1}{w_{i+1}} - p_i, -\frac{1}{w_i} - p_i) \cap [0, +\infty)$.

Proof: From (34)-(35), $g'_h(p_i^+)$ and $g'_h(p_i^-)$ are functions of $\frac{1}{h}$ for a given p_i . If $(-\frac{1}{w_{i+1}} - p_i, -\frac{1}{w_i} - p_i) \cap [0, +\infty)$ is not empty, it is easy to verify that $0 = g'_h(p_i^-) > g'_h(p_i^+)$ when $\frac{1}{h} = -\frac{1}{w_i} - p_i$, and $g'_h(p_i^-) > g'_h(p_i^+)$ and $g'_{h+}(p_i) \leq 0$ when $\frac{1}{h} = -\frac{1}{w_{i+1}} - p_i$. Since given p_i , $g'_h(p_i^-)$ and $g'_h(p_i^+)$ increase as $\frac{1}{h}$ decreases, then decreasing $\frac{1}{h}$ from $-\frac{1}{w_i} - p_i$ to $\max\{0, -\frac{1}{w_{i+1}} - p_i\}$, we have $g'_h(p_i^-) \geq 0 \geq g'_h(p_i^+)$ for all $\frac{1}{h} \in (-\frac{1}{w_{i+1}} - p_i, -\frac{1}{w_i} - p_i) \cap [0, +\infty)$. By Lemma 2, the proposition follows. ■

Propositions 5 and 6 obtain the optimal solution for $\frac{1}{h} \in [-\frac{1}{w_N} - p_N, -\frac{1}{w_1}] \cap [0, \infty)$. For other $h \geq 0$, using Conditions 2 and 3 in Lemma 2, we can prove the optimality of (36) as follows.

- Proposition 7:* 1) For any h such that $\frac{1}{h} \geq -\frac{1}{w_1}$, the optimal solution is $p^*(h) = 0$;
2) For any h such that $0 \leq \frac{1}{h} \leq -\frac{1}{w_N} - p_N$, the optimal solution $p^*(h) = p_N$.

Proof: Note that, since $U(B, p, 1)$ is non-increasing with respect to p , we have $-\frac{1}{w_1} \geq 0$. When $\frac{1}{h} = -\frac{1}{w_1}$, it is easy to verify that $g'_h(0^+) = 0$. Since $g'_h(0^+)$ is also a function of $\frac{1}{h}$ which decreases as $\frac{1}{h}$ increases, we have for any $\frac{1}{h} \geq -\frac{1}{w_1}$, $g'_h(0^+) \leq 0$. By Condition 3 in Lemma 2, we must have $p^*(h) = 0$ for any h such that $\frac{1}{h} \geq -\frac{1}{w_1} \geq 0$. Similarly, we may also verify that for any $\frac{1}{h} \leq -\frac{1}{w_N} - p_N$, $g'_h(p_N^-) \geq 0$. By Condition 2 in Lemma 2, we must have $p^*(h) = p_N$ for any h such that $0 \leq \frac{1}{h} \leq -\frac{1}{w_N} - p_N$. ■

Note that, given $B \in \mathcal{B}_\delta$, $p^*(h)$ is a piecewise function in closed-form. Then, $V^k(B, A)$ can be efficiently evaluated as

$$V^k(B, 1) = \mathbb{E}_{h_k} \left\{ \log(1 + p^*(h_k)h_k) + \gamma U^k(B, p^*(h_k), 1) \right\}, \quad (38)$$

and

$$V^k(B, 0) = \mathbb{E}_{h_k} \left\{ \log(1) + \gamma U^k(B, 0, 0) \right\} = U^k(B, 0, 0), \quad (39)$$

where

$$U^k(B, p, 0) = \sum_{A'=0,1} \Pr(A_{k+1} = A' \mid A_k = 0) [W_\delta^{k+1}(\min\{b_{\max}, B + e_k - pT_c\}, A')]. \quad (40)$$

B. Calculating the Approximate Value Function

In order to obtain the power allocation, we need to compute the approximate value function given by (13)-(15) for $k = 1, 2, \dots, K$ ($K = \infty$ for infinite-horizon case). Then, when the observation is available,

we solve the problem given in (18).

1) *Finite K*: We first consider the finite-horizon case where K is finite, we assume that the distributions of channel fading are independent across slots but not necessarily identical.

The power allocation consists of two phases. In the first phase, we recursively compute the approximate value function from $k = K$ to $k = 1$, following (13)-(15). Specifically, in the i -th iteration, we obtain $W_\delta^{(i)}(b, A)$ for slot $k = K - i + 1$ as follows. Based on $W_\delta^{(i-1)}(b, A)$ obtained in the previous iteration (or the initial function for the first iteration), for each $B \in \mathcal{B}_\delta$ and $A = \{0, 1\}$, we obtain the piecewise linear function $U^{(i)}(B, p, A)$ by specifying the sets \mathcal{I} and \mathcal{W} . Then, we use (36) to obtain $p^*(h)$ and use (38)-(39) to update $V^{(i)}(B, A)$ for all $B \in \mathcal{B}_\delta$ and $A = \{0, 1\}$. With the set $\{V^{(i)}(B, A) \mid B \in \mathcal{B}_\delta, A = \{0, 1\}\}$, the approximate value function $W_\delta^{(i)}(b, A)$ can be obtained using (15) and we store the closed-form $W_\delta^{(i)}(b, A)$ in a look-up table. Note that the above first phase should be completed before the first slot.

The second phase is performed at the beginning of each slot, once the observation becomes available. This phase is to solve the problem given in (18) using (36). Specifically, at the beginning of slot k , the transmitter observes the system state, i.e., the channel access state A , the channel gain h , and the current battery state b . When $A = 0$, the transmitter keeps silent. Otherwise, the transmitter retrieves the approximate value function $W_\delta^{k+1}(b, A)$ (i.e., $W_\delta^{(K-i-1)}(b, A)$) from the look-up table and then calculate the power allocation using (36).

The entire computational procedure for the finite-horizon case is summarized in Algorithm 1.

Algorithm 1 - Finite-Horizon Power Allocation

1: Inputs
Distributions of \mathbf{H} , \mathbf{A} ; value of e_k for $k = 1, 2, \dots, K$
The approximation precision $\delta > 0$ and the discount factor $\gamma \in [0, 1]$.

2: Phase-I: Compute the approximate value function update (offline calculation)
FOR $k = K$ **TO** 1
(*) Calculate $V^k(B, h, A)$ for $B \in \mathcal{B}_\delta, A \in \{0, 1\}$ using (36) and (38)-(39)
Compute $W_\delta^k(b, A)$ from $V^k(B, A)$ using (15) and store it
ENDFOR

3: Phase-II: Power Allocation (online calculation)
FOR $k = 1$ **TO** K
Get the observations $\Gamma_k = (b_k, h_k, A_k)$
Retrieve $W_\delta^{k+1}(b, A)$ and calculate $U^k(b_k, p, A_k)$ using (13)
Calculate $p^*(h_k)$ using (36)
ENDFOR

Remark 1: If the observations can be predicted in a scheduling period K , i.e., \mathbf{H} , \mathbf{E} , and \mathbf{A} are known in advance, we can rewrite (5) as follows

$$\mathcal{P}^* = \arg \max_{p_k, k=1, 2, \dots, K} \left\{ \sum_{k=1}^K A_k \log(1 + p_k h_k) \right\}, \quad (41)$$

subject to the constraints in (2), (48), and (4) for $k = 1, 2, \dots, K$.

We note that in the above case all the observations are non-causally known in advance and the problem in (41) is a convex optimization problem. Instead of the generic convex solver, there is also an efficient dynamic water-filling algorithm proposed in [9], for solving (41) optimally. Moreover, since (41) is a special case of the stochastic case, Algorithm 1 is also applicable and would approach the optimal performance as the dynamic water-filling algorithm when $\delta \rightarrow 0$. Specifically, the use of Algorithm 1 or the dynamic water-filling algorithm strikes a balance between the performance and the computational complexity.

2) *Infinite K :* In the infinite-horizon case, although K is infinite, the number of the iterations in the first phase is not infinite since the approximate value function will converge. Moreover, since we have assumed that e_k is static and h_k is i.i.d., the converged approximate value function can be directly used in (18) to obtain the power allocation with the observations in the second phase, for all slots.

We denote

$$\mathcal{T}_\delta : W_\delta(b, A) \rightarrow W_\delta(b, A) \quad (42)$$

as the value function update operator in (13)-(15): based on a given value function $W_\delta^{(i)}(b, A)$, it solves (14) to obtain $V^{(i)}(B, p, A)$ for $B \in \mathcal{B}_\delta$, and then generates the new approximate value function $W_\delta^{(i+1)}(b, A)$ by (15). Then we can write

$$W_\delta^{(i+1)}(b, A) \triangleq \mathcal{T}_\delta \left[W_\delta^{(i)}(b, A) \right], \quad b \in [0, b_{\max}]. \quad (43)$$

Note that \mathcal{T}_0 is the standard Bellman operator corresponding to (6)-(7) without the value function approximation, i.e., $\delta = 0$ [15].

Then the computational procedure for the infinite-horizon case is summarized in Algorithm 2.

Algorithm 2 - Infinite-Horizon Power Allocation

- 1: Inputs
 Distributions of h, A ; value of e
 The approximation precision $\delta > 0$, the discount factor $\gamma \in (0, 1)$, and the termination condition α .
 - 2: Phase-I Approximate value function update (offline calculation)
 $i \leftarrow 0$
REPEAT
 (*) $W_\delta^{(i+1)}(b, A) = \mathcal{T}_\delta \left[W_\delta^{(i)}(b, A) \right]$
 $i \leftarrow i + 1$
UNTIL $\|W_\delta^{(i)}(b, A) - W_\delta^{(i-1)}(b, A)\|_\infty \leq \alpha$
 $W_\delta^*(b, A) \leftarrow W_\delta^{(i)}(b, A)$
 - 3: Phase-II Power Allocation (online calculation)
AT THE BEGINNING OF EACH SLOT
 Get the observations $\Gamma = (b, h, A)$
 Retrieve $W_\delta^*(b, A)$ and calculate $U^*(b, p, A)$ using (13)
 Calculate $p^*(h)$ using (36)
-

To show the convergence of the approximate value function update, we first note that, by repeatedly performing \mathcal{T}_0 on any initial value function, a converged value function can be obtained as follows [15]:

$$v^*(b, A) \triangleq \mathcal{T}_0 \cdot \mathcal{T}_0 \cdot \dots \left[v^{(1)}(b, A) \right] = \mathcal{T}_0^\infty \left[v^{(1)}(b, A) \right]. \quad (44)$$

Extending the convergence of \mathcal{T}_0 to \mathcal{T}_δ , we introduce the following lemma. The proof is given in Appendix A.

Lemma 3: The operator \mathcal{T}_δ has the γ -contraction property, i.e., for any two functions $V_1(b, A)$ and $V_2(b, A)$, we have

$$\|\mathcal{T}_\delta[V_1(b, A)] - \mathcal{T}_\delta[V_2(b, A)]\|_\infty \leq \gamma \|V_1(b, A) - V_2(b, A)\|_\infty. \quad (45)$$

It then follows that

$$\|\mathcal{T}_\delta^{i+1}[W_\delta^{(1)}(b, A)] - \mathcal{T}_\delta^i[W_\delta^{(1)}(b, A)]\|_\infty \leq \gamma^i \|\mathcal{T}_\delta[W_\delta^{(1)}(b, A)] - W_\delta^{(1)}(b, A)\|_\infty, \quad (46)$$

i.e., $\mathcal{T}_\delta^i[W_\delta^{(1)}(b, A)]$ converges as i increases. Moreover, the error between the converged approximate value function and $v^*(b, A)$ is bounded as follows.

Theorem 3: If $\|\mathcal{T}_\delta^i[W_\delta^{(1)}(b, A)] - \mathcal{T}_\delta^{i-1}[W_\delta^{(1)}(b, A)]\|_\infty \leq \alpha$, then the error between $v^*(b, A)$ and $W_\delta^{(i)}(b, A)$ is bounded by

$$\|W_\delta^{(i)}(b, A) - v^*(b, A)\|_\infty \leq \frac{\gamma\alpha + \|2v^*(\delta, A) - v(0, A) - v^*(2\delta, A)\|_\infty}{1 - \gamma}. \quad (47)$$

Proof: The proof is provided in Appendix B. ■

Note that, Algorithms 1 and 2 have both the offline calculation part and the online calculation part. During offline calculation, we evaluate $V^k(B, A)$ for each $B \in \mathcal{B}_\delta$ in each iteration, i.e., solve $\mathcal{O}(B_{\max}/\delta)$ convex optimization problems in each iteration. Specifically, rather than using an exhaustive search for each combination of the discretized (B, H) (H is the discretized channel gain) as done by the standard discrete MDP method, the proposed algorithms use (36) to calculate $V^k(B, A)$ for each $B \in \mathcal{B}_\delta$ directly. Moreover, for the infinite case, by Lemma 3, the α -converged approximate value function can be obtained within $\mathcal{O}(\log_\gamma \alpha)$ iterations. On the other hand, during online calculation, we retrieve $W_\delta^{k+1}(b, A)$ (or $W_\delta^*(b, A)$) from the look-up table and then use (36) to compute the power allocation for the specific observation (b_k, h_k, A_k) .

Moreover, the proposed algorithms calculate the power allocation based on the continuous battery state and channel gain, and the obtained power allocation is also continuous. Thus it provides higher precision for both offline calculation and online calculation than the conventional discrete MDP method, especially when the discretization step is large. Finally, as shown in Section VI, a better performance can be achieved by the proposed algorithm with a lower computational complexity compared with the conventional discrete

MDP method.

V. POWER ALLOCATION WITH IMPERFECT ENERGY PREDICTION

Although energy harvesting is usually predictable, there may exist a non-negligible prediction error in practice. In this section, we treat the case of imperfect energy harvesting prediction where the prediction error is an i.i.d. random variable. We also consider a general payoff function $r(p, A)$, which is continuous, non-decreasing and concave with respect to p given $A \in \{0, 1\}$.

In this general model, we assume the energy harvesting process consisting of a deterministic part e_k and a stochastic part ε_k . The deterministic process e_k in practice is obtained from the prediction using historic observations, e.g., by averaging the historic measurement with the weather adjustment.

With the prediction error, the problem formulation is modified as follows:

$$b_{k+1} = \min \{b_{\max}, b_k + e_k + \varepsilon_k - p_k T_c\} , \quad (48)$$

and

$$\mathcal{P}^* \triangleq \arg \max_{p_k(\cdot), k=1,2,\dots,K} \left\{ \mathbb{E}_{\mathbf{H}, \mathbf{E}, \mathbf{A}} \left[\sum_{k=1}^K \gamma^{k-1} r(p_k(\Gamma_k), h_k) \right] \right\} , \quad (49)$$

subject to the constraints in (2), (4), and (48), for $k = 1, 2, \dots, K$, where $\mathbf{E} \triangleq [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_K]$. Accordingly, since ε_k is a random variable, the (approximate) value function update rules in (7) and (13) are changed to

$$u^k(b_k, p_k, A_k) \triangleq \mathbb{E}_{\varepsilon_k, A_{k+1}|A_k} [v^{k+1}(\min\{b_{\max}, b_k + e_k + \varepsilon_k - p_k T_c\}, A_{k+1})] , \quad (50)$$

and

$$U^k(b_k, p_k, A_k) \triangleq \mathbb{E}_{\varepsilon_k, A_{k+1}|A_k} [W_\delta^{k+1}(\min\{b_{\max}, b_k + e_k + \varepsilon_k - p_k T_c\}, A_{k+1})] , \quad (51)$$

respectively.

Obviously, since $r(p, A)$ is continuous, non-decreasing and concave with respect to p given A , and the expectation with respect to ε_k in (50) and (51) preserves the concavity and the non-decreasing properties, we can extend the analysis in Section III to the case with the general payoff function and imperfect energy prediction, obtaining the same concavity and non-decreasing properties.

However, note that, the optimal solution $p^*(h)$ in (36) is based on the facts that $U^k(b, p, A)$ is a piecewise linear function and $r(p, h) = \log(1 + ph)$, which are no longer valid with the general payoff function

and/or imperfect energy prediction. Then, in Algorithms 1 and 2, the steps marked by (*), which aim to solve the problem in (14), need to be modified accordingly. In particular, we now need to use some standard convex solver to numerically solve (14).

VI. SIMULATION RESULTS

We use the payoff function $r(p, h) = \log(1 + ph)$. We assume that the channel fading h_k is an i.i.d. random variable following the Rayleigh distribution with the parameter σ . We first assume that the harvested energy can be perfectly predicted. For the transmitter, we set the maximum transmission power as $p_{\max} = 6$ units per slot, the battery capacity as $b_{\max} = 15$ units, and the initial battery level as $b_0 = 2$ units. Further, we set the probability of the channel access suspension as $q = \hat{q} = 0.1$, the approximate precision δ of the approximating value function as 1 and 0.1, and the convergence error tolerance for the infinite-horizon case as $\alpha = 0.0001$.

We first evaluate the performance of the proposed algorithms. For comparison, we consider three simple power allocation methods, the *greedy policy*, the *balanced policy*, and the standard discrete MDP method. The greedy policy tries to allocate as much power as possible in each slot subject to the energy availability. On the other hand, the balanced policy tries to allocate a constant power in each slot, e.g., the mean value of the harvested energy. Moreover, for the standard discrete MDP method, we discretize the battery level, the channel gain, and the transmission power with the same precision factor δ , and then perform the dynamic programming algorithm and the value iteration algorithm on the discrete state space for the finite- and infinite-horizon cases, respectively.

For the finite-horizon case, we set $K = 30$, $\gamma = 1$, and $\sigma = 0.7, 0.8, 0.9, 1.0, 1.1, 1.2$. We randomly generate the prediction value e_k following a positive truncated-Gaussian distribution with the variance of 2. We consider two typical scenarios, an *energy-constrained scenario* with the mean of the harvested energy of 2, and a *power-constrained scenario* with the mean of the harvested energy of 4. In the energy-constrained scenario, the average harvested energy is much lower than the maximum transmission power and the energy schedule is mainly constrained by the energy availability. On the other hand, in the power-constrained scenario, the average harvested energy approaches to the maximum transmission power and this constraint dominates the energy scheduling. For both scenarios, we compare the performance of the proposed algorithm with the standard discrete MDP method, the greedy policy and the balanced policy, averaged over 2×10^6 realizations in Fig. 6 and Fig. 7, respectively. Although we cannot obtain the optimal performance, we utilize the error bound given in (16) and (27) as an upper-bound of the optimal

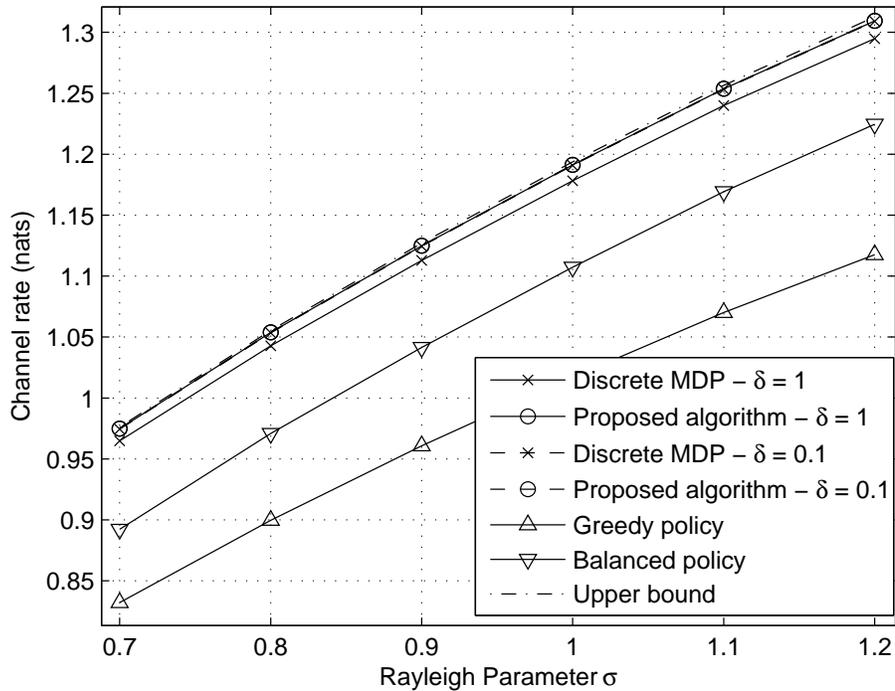


Fig. 6. Performance comparisons in the energy-constrained scenario for the finite-horizon case.

performance. Also, the performance obtained by the standard discrete MDP method can serve as the lower-bound.

It can be seen from Fig. 6 and Fig. 7 that for $\delta = 1$, the performance of the proposed algorithm tightly approaches the upper-bound of the optimal performance in both scenarios while there is a gap between the proposed algorithm and the standard discrete MDP method. It is mainly because that the discrete MDP method discretizes all continuous variables and causes some non-negligible error with the large discretization step. For $\delta = 0.1$, both the proposed algorithm and the standard discrete MDP method achieve the comparable performance, but their computational complexities are not comparable, e.g., the exhaustive search is involved in the latter. The greedy and balanced policies both have significantly inferior performances. Moreover, we note that the total rate increases as the Rayleigh parameter σ increases and the rate in the energy-constrained scenario is higher than that in the power-constrained scenario.

For the infinite-horizon case, we set $\gamma = 0.85$, $e_k = 3$, and $\sigma = 0.7, 0.8, 0.9, 1.0, 1.1, 1.2$. Similar to the finite-horizon case, we evaluate the performance for various power allocation policies, averaged over 2×10^6 realizations. The performance comparisons for various power allocation policies are shown in Fig. 8. Moreover, the convergence behavior of the proposed algorithm is also shown in Fig. 9 for $\sigma = 1$ and $\gamma = 0.8, 0.85, 0.9$.

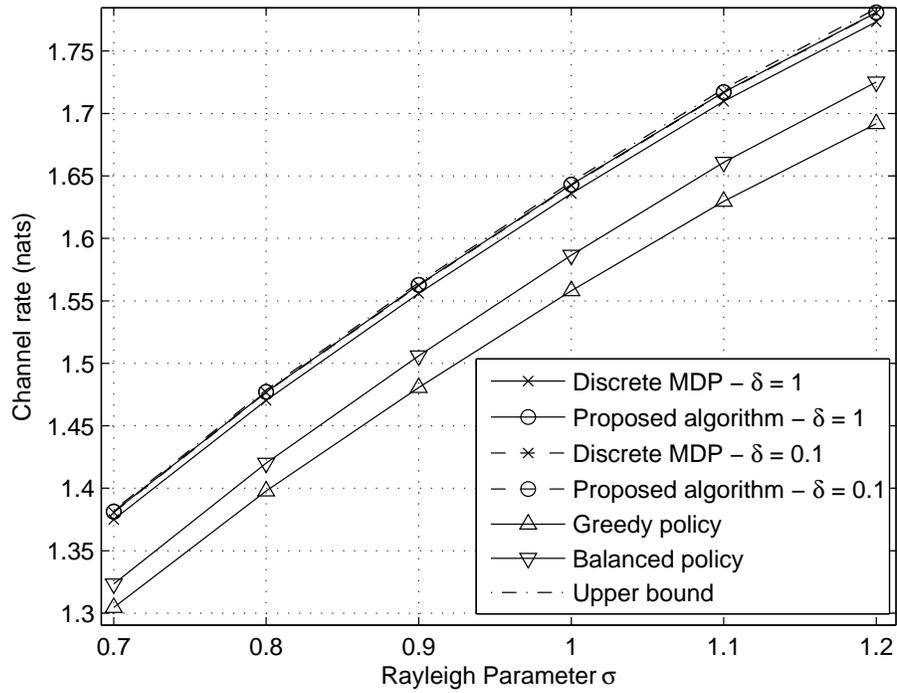


Fig. 7. Performance comparisons in the power-constrained scenario for the finite-horizon case.

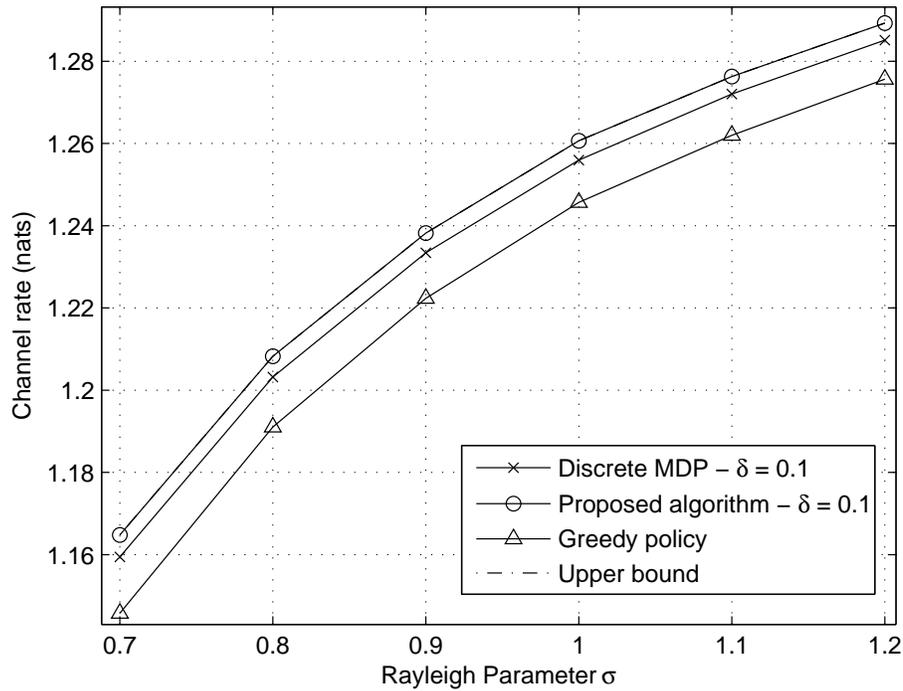


Fig. 8. Performance comparisons for the infinite-horizon case.

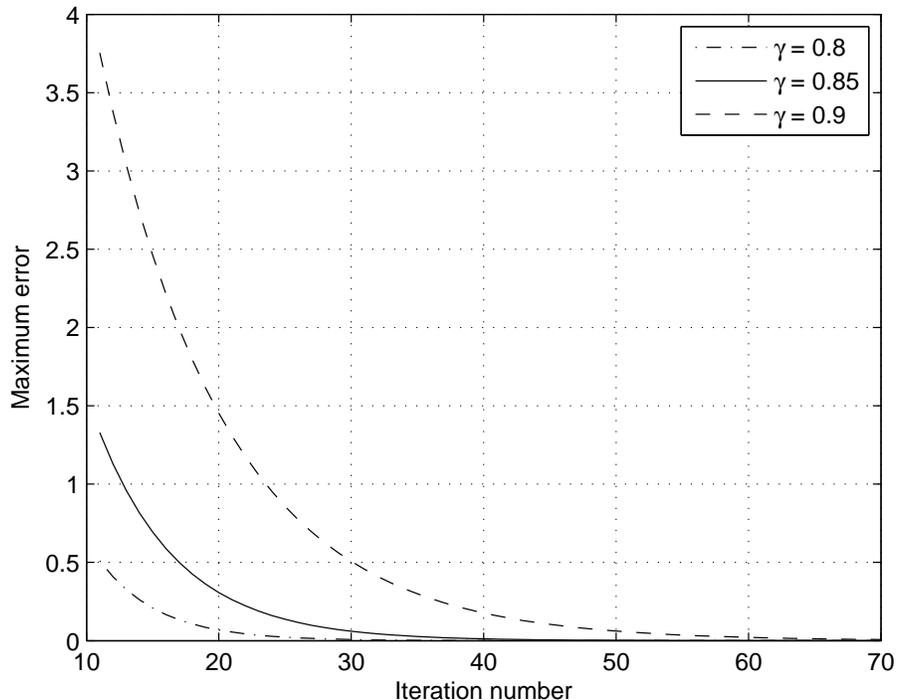


Fig. 9. The convergence of Algorithm 2 for $\sigma = 1$.

Similar to the finite-horizon case, it is seen from Fig. 8 that the proposed algorithm has the best performance, tightly approaching the upper-bound of the optimal performance. We note that the standard discrete MDP method with a discretization step of $\delta = 0.1$ performs worse than the proposed algorithm. Further, the approximation gap is slightly higher in the infinite case as compared to that in the finite case. Moreover, we see that the greedy approach has the worst performance. In addition, it is seen from Fig. 9 that the discount factor affects the convergence speed, as analyzed in Section IV. Also, in the simulations for $\gamma = 0.8, 0.85, 0.9$, the proposed algorithm converges within around 30, 40 and 70 iterations, respectively.

We next evaluate the impact of the imperfect prediction error. We consider the finite-horizon case and set $K = 10$, $\gamma = 1$, $e_k = 3.5$, $\sigma = 1$, $q = 1 - \hat{q} = 0$, and $\delta = 0.1$. In this scenario, we only consider the impact of the imperfect prediction and we assume that the channel fading is known and the energy prediction error follows the discrete uniform distribution between $-v$ and v with the step of 0.1. The total payoff obtained by the proposed algorithm with causal information and the water-filling based algorithm in [9] with non-causal information is compared in Fig. 10, over different prediction error ranges $v = 0, 0.5, 1, 1.5, 2, 2.5$. It is seen from Fig. 10 that as v decreases, the performance gap of the two algorithms with and without non-causal information decreases and approaches zero.

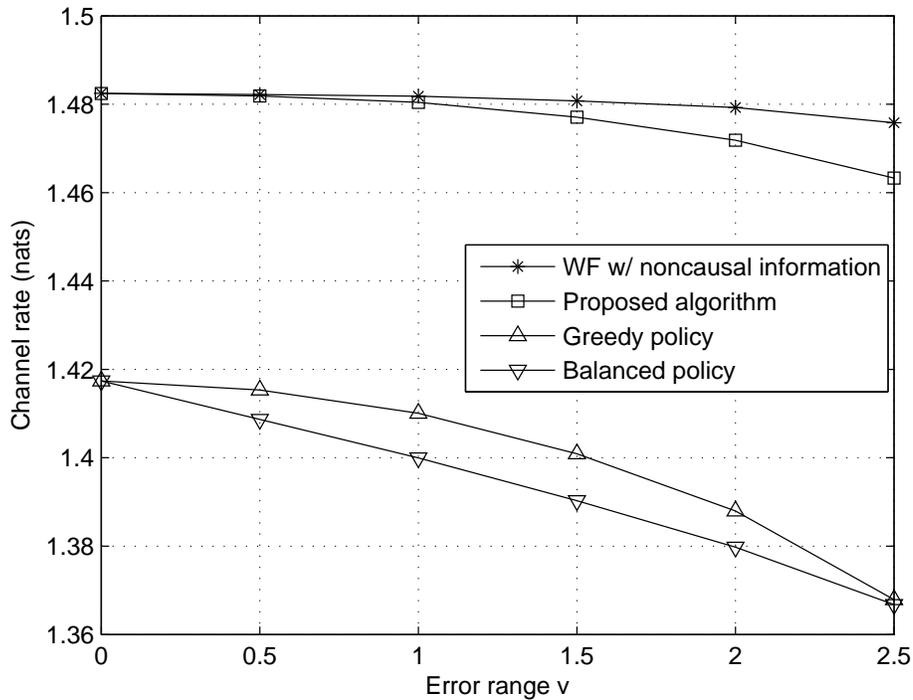


Fig. 10. Performance comparisons for the finite-horizon case with different prediction error ranges.

VII. CONCLUSIONS

We have considered the problem of optimal power allocation for an access-controlled transmitter with energy harvesting capability, operating in time-slotted fashion with causal knowledge of the channel state and the energy harvesting state. The energy harvesting process is a sum of a deterministic non-causal estimate and a random causal prediction error. This problem is formulated as a Markov decision process with continuous state. To efficiently solve this problem for both the finite- and infinite-horizon cases, we have introduced the approximate value function and developed efficient algorithms for obtaining the approximately optimal solutions. The proposed algorithms provide an approximately optimal continuous power allocation, whose performance is better than that obtained by the standard discrete MDP method, in a computationally efficient manner. Simulation results demonstrate that the proposed algorithms can closely approach the optimal performance for both the finite- and infinite-horizon cases.

APPENDIX A

PROOF OF LEMMA 3

It is known that \mathcal{T}_0 , which is the operator in the standard value iteration algorithm, is a γ -contraction [15]. Denoting $(b^*, A^*) \triangleq \arg \|\mathcal{T}_0[V_1(b, A)] - \mathcal{T}_0[V_2(b, A)]\|_\infty$, for any (B_0, A_0) and $(B_0 + \delta, A_0)$ where

$B_0, B_0 + \delta \in \mathcal{B}_\delta, A_0 \in \{0, 1\}$, we have that

$$\left| \mathcal{T}_0[V_1](B_0, A_0) - \mathcal{T}_0[V_2](B_0, A_0) \right| \leq \left| \left(\mathcal{T}_0[V_1] - \mathcal{T}_0[V_2] \right)(b^*, A^*) \right| \quad (52)$$

and

$$\left| \mathcal{T}_0[V_1](B_0 + \delta, A_0) - \mathcal{T}_0[V_2](B_0 + \delta, A_0) \right| \leq \left| \left(\mathcal{T}_0[V_1] - \mathcal{T}_0[V_2] \right)(b^*, A^*) \right|. \quad (53)$$

Note that, given a value function $V(b, A)$, $\mathcal{T}_\delta[V](b, A)$ is the piecewise linear function reconstructed from the sample set $\{\mathcal{T}_0[V](B, A) \mid B \in \mathcal{B}_\delta\}$, as in (15). Since $B_0, B_0 + \delta \in \mathcal{B}_\delta$, then for any $b \in [B_0, B_0 + \delta]$, we have

$$\begin{aligned} & \left| \mathcal{T}_\delta[V_1](b, A) - \mathcal{T}_\delta[V_2](b, A) \right| \\ & \leq \max \left\{ \left| \mathcal{T}_\delta[V_1](B_0, A) - \mathcal{T}_\delta[V_2](B_0, A) \right|, \left| \mathcal{T}_\delta[V_1](B_0 + \delta, A) - \mathcal{T}_\delta[V_2](B_0 + \delta, A) \right| \right\} \\ & = \max \left\{ \left| \mathcal{T}_0[V_1](B_0, A) - \mathcal{T}_0[V_2](B_0, A) \right|, \left| \mathcal{T}_0[V_1](B_0 + \delta, A) - \mathcal{T}_0[V_2](B_0 + \delta, A) \right| \right\} \end{aligned} \quad (54)$$

Since B_0 and A are arbitrarily chosen from $\mathcal{B}_\delta / \max\{\mathcal{B}_\delta\}$ and $\{0, 1\}$, respectively, we have

$$\begin{aligned} & \left\| \mathcal{T}_\delta[V_1(b, A)] - \mathcal{T}_\delta[V_2(b, A)] \right\| \\ & \leq \max \left\{ \left| \mathcal{T}_0[V_1](B_0, A_0) - \mathcal{T}_0[V_2](B_0, A_0) \right|, \left| \mathcal{T}_0[V_1](B_0 + \delta, A_0) - \mathcal{T}_0[V_2](B_0 + \delta, A_0) \right| \right\} \end{aligned} \quad (55)$$

$$\leq \left| \left(\mathcal{T}_0[V_1] - \mathcal{T}_0[V_2] \right)(b^*, A^*) \right| \quad (56)$$

$$= \left\| \mathcal{T}_\delta[V_1(b, A)] - \mathcal{T}_\delta[V_2(b, A)] \right\|_\infty \quad (57)$$

$$\leq \gamma \left\| V_1(b, A) - V_2(b, A) \right\|_\infty \quad (58)$$

where (55) follows from (54), (56) follows from (52)-(53), and (57) follows the definition of (b^*, A^*) .

APPENDIX B

PROOF OF THEOREM 3

Denote $\beta(b, A) \triangleq v^*(b, A) - \mathcal{T}_\delta[v^*(b, A)]$. By Lemma 3, we have

$$\begin{aligned} & \left\| W_\delta^{(i)}(b, A) - v^*(b, A) \right\|_\infty \\ & = \left\| W_\delta^{(i)}(b, A) + W_\delta^{(i+1)}(b, A) - W_\delta^{(i+1)}(b, A) - v^*(b, A) \right\|_\infty \end{aligned}$$

$$\begin{aligned}
&\leq \|W_\delta^{(i)}(b, A) - W_\delta^{(i+1)}(b, A)\|_\infty + \|W_\delta^{(i+1)}(b, A) - v^*(b, A)\|_\infty \\
&= \|\mathcal{T}_\delta[W_\delta^{(i)}(b, A)] - \mathcal{T}_\delta[W_\delta^{(i-1)}(b, A)]\|_\infty + \|\mathcal{T}_\delta[W_\delta^{(i)}(b, A)] - \mathcal{T}_\delta[v^*(b, A)] - \beta(b, A)\|_\infty \\
&= \|\mathcal{T}_\delta[W_\delta^{(i)}(b, A)] - \mathcal{T}_\delta[W_\delta^{(i-1)}(b, A)]\|_\infty + \|\mathcal{T}_\delta[W_\delta^{(i)}(b, A)] - \mathcal{T}_\delta[v^*(b, A)]\|_\infty + \|\beta(b, A)\|_\infty \\
&\leq \gamma \|W_\delta^{(i)}(b, A) - W_\delta^{(i-1)}(b, A)\|_\infty + \gamma \|W_\delta^{(i)}(b, A) + v^*(b, A)\|_\infty + \|\beta(b, A)\|_\infty
\end{aligned} \tag{59}$$

where (59) follows the γ -contraction of the operator \mathcal{T}_δ .

From (59), we have

$$\begin{aligned}
\|W_\delta^{(i)}(b, A) - v^*(b, A)\|_\infty &\leq \frac{\gamma \|W_\delta^{(i)}(b, A) - W_\delta^{(i-1)}(b, A)\|_\infty + \|\beta(b, A)\|_\infty}{1 - \gamma} \\
&\leq \frac{\gamma\alpha + \|\beta(b, A)\|_\infty}{1 - \gamma}
\end{aligned} \tag{60}$$

Also, since the only difference between \mathcal{T}_δ and \mathcal{T}_0 is the approximation process, then we have $\beta(b, A) = v^*(b, A) - \mathcal{T}_\delta[v^*(b, A)] = v^*(b, A) - \mathcal{L}[\mathcal{T}_0[v^*(b, A)], \delta] = v^*(b, A) - \mathcal{L}[v^*(b, A), \delta]$. Using Proposition 4, we have

$$\|\beta(b, A)\|_\infty \leq \|2v^*(\delta, A) - v(0, A) - v^*(2\delta, A)\|_\infty \leq \|v^*(\delta, A) - v^*(0, A)\|_\infty. \tag{61}$$

Therefore, (60) can be further written as

$$\|W_\delta^{(i)}(b) - v^*(b)\|_\infty \leq \frac{\gamma\alpha + \|2v^*(\delta, A) - v(0, A) - v^*(2\delta, A)\|_\infty}{1 - \gamma}. \tag{62}$$

REFERENCES

- [1] C. Han and et al, "Green radio: radio techniques to enable energy-efficient wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 46–54, Jun. 2011.
- [2] Y. Chen and et al, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, Jun. 2011.
- [3] S. Yeh, "Green 4G communications: renewable-energy-based architectures and protocols," in *Proc. 2010 Global Mobile Congress*, Oct. 2010, pp. 1–5.
- [4] S. Sudevalayam and P. Kulkarni, "Energy harvesting sensor nodes: survey and implications," *IEEE Commun. Surveys Tuts.*, vol. 13, no. 3, pp. 443–461, Sep. 2011.
- [5] A. P. Bianzino and et al, "A survey of green networking research," *IEEE Commun. Surveys Tuts.*, vol. 14, no. 1, pp. 3–20, Jan. 2012.
- [6] S. Chen, P. Sinha, N. Shroff, and C. Joo, "Finite-horizon energy allocation and routing scheme in rechargeable sensor networks," in *Proc. IEEE 2011 INFOCOM*, Apr. 2011, pp. 2273–2281.
- [7] C. Ho and R. Zhang, "Optimal energy allocation for wireless communications with energy harvesting constraints," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4808–4818, Sep. 2012.

- [8] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.
- [9] Z. Wang, V. Aggarwal, and X. Wang, "Iterative dynamic water-filing for fading multiple-access channels with energy harvesting," *IEEE Trans. Signal Process.*, submitted for publication.
- [10] L. Huang and M. Neely, "Utility optimal scheduling in energy-harvesting networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1117–1130, Aug. 2013.
- [11] P. Blasco, D. Gunduz, and M. Dohler, "A learning theoretic approach to energy harvesting communication system optimization," *IEEE Trans. Wireless Commun.*, vol. 12, no. 4, pp. 1872–1882, Apr. 2013.
- [12] Q. Bai, R. Amjad, and J. Nosssek, "Average throughput maximization for energy harvesting transmitters with causal energy arrival information," in *Proc. IEEE 2013 WCNC*, Apr. 2013, pp. 4232–4237.
- [13] J. Piorno, C. Bergonzini, K. Atienza, and T. Rosing, "Prediction and management in energy harvested wireless sensor nodes," in *Proc. VITAE 2009*, May 2009, pp. 6–10.
- [14] J. Lu, S. Liu, Q. Wu, and Q. Qiu, "Accurate modeling and prediction of energy availability in energy harvesting real-time embedded systems," in *Proc. Green Computing Conf. 2010*, Aug. 2010, pp. 469–476.
- [15] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. New York: John Wiley & Sons, 1994.
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge: Cambridge University Press, 2009.